

Developing and assessing prediction intervals for soil property maps derived from legacy databases

J. L. Helmick

Department of Statistics, West Virginia University, Morgantown, West Virginia, USA

T. W. Nauman & J. A. Thompson

Division of Plant and Soil Sciences, West Virginia University, Morgantown, West Virginia, USA

ABSTRACT: The *GlobalSoilMap* project aims to create a global grid of a variety of soil functional properties at a fine resolution. Uncertainty surrounding these property estimates is of utmost importance when utilizing soil maps for predictive purposes. For the initial version of the map being produced of the United States, property values were estimated from the U.S. General Soil Map (STATSGO2) database, which is a broad-based inventory of soil data recorded across the United States. Multiple soil and non-soil (water, rock outcrop, urban areas, etc.) components are aggregated in STATSGO2 into groups of similar polygons, called map units. Each soil component includes various estimated property values for each horizon associated with that specific soil (if applicable), including an estimated upper, lower, and representative value. Hierarchically, STATSGO2 includes multiple horizon values for each component, which make up the map units. Previous work has used area-weighted means of the representative values for each component to develop a representative value for the map unit. For this study, prediction intervals were developed from the low and high estimated property values provided in the database. Instead of calculating a weighted mean of the low and high estimated property values, the lower prediction limit was determined as the lowest of the low values associated with any of the components in the map unit. The upper prediction limit was determined as the highest of the high values in the same manner. For each map unit, this method provided a unique prediction interval which was likely to encompass property values of soil typically found in that map unit. We empirically evaluated the soil property prediction intervals derived from STATSGO2 for three soil properties: organic carbon content, pH, and clay content. Using measured property data from up to 722 pedons from the National Cooperative Soil Survey database, prediction intervals were assessed by modeling their coverage accuracy over a set of external validation data. The effects of soil depth, soil order, temperature regime, and moisture regime on prediction interval coverage were analyzed, and coverage was found to be 87.6% for organic carbon, 90.6% for pH, and 86.4% for clay. It is shown that legacy data from the United States that includes low and high property methods can be used to represent uncertainty in the form of prediction intervals. Coverage based on these methods is only slightly below the nominal level of 95%. Consistency of these intervals was demonstrated across a variety of soil orders, temperature regimes, and moisture regimes.

1 INTRODUCTION

The *GlobalSoilMap* project aims to create a global grid of a variety of soil functional properties at a fine resolution (3 arc-second by 3 arc-second). These digital products are designed to provide estimates for a minimum data set of 12 soil properties including depth to rock, plant exploitable (effective) depth, organic carbon, pH, clay, silt, sand, coarse fragments, effective cation exchange capacity (ECEC), bulk density (whole soil), bulk density (fine earth), and available water capacity (GlobalSoilMap Science Committee 2013). In addition to point estimates for these values, uncertainty of these estimates are to be presented as 95% prediction intervals (PIs) about the point estimate (GlobalSoilMap Science Committee

2013). Estimates for all properties except depth to rock and effective depth are made for six standard depth increments (0-5, 5-15, 15-30, 30-60, 60-100, and 100-200 cm).

For the initial version of the map being produced of the conterminous United States, property values were estimated from the US General Soil Map (STATSGO2) database (USDA-NRCS 2011), which is a broad-based inventory of soil data recorded across the United States. The STATSGO2 database aggregates soil information into polygons that are attributed with multiple components. Components may be soil (series or higher-level taxa) or non-soil (water, rock outcrop, urban areas, etc.). Tables within the relational database include: (i) the map unit table, which lists the spatial map units; (ii) the com-

ponent table, which includes the proportion of up to 21 components in each map unit; and (iii) the horizon table, which lists the number and type of horizons associated with each component as well as their properties. STATSGO2 reports an estimated upper, lower, and representative value for a suite of soil properties. For each component, an equal-area spline function was applied to the horizon data to calculate soil property values for each of six standard soil depth increments prescribed by *GlobalSoilMap*. The component percentages were then used to calculate an area-weighted mean of the representative values for each component to develop a representative value for the map unit. The procedures for developing these initial *GlobalSoilMap* products for the conterminous US are described in greater detail by Odgers et al. (2012).

Uncertainty in soil property estimates is of great interest to the *GlobalSoilMap* project and the broader digital soil mapping community. Here we describe an approach for obtaining reasonable PIs that are likely to contain the true property values of soils of the conterminous US, and which are necessary for interpretation and use of digital soil maps. Whether used for modeling purposes or by scientists in the field, maps should accurately portray the uncertainty in soil property estimates.

2 METHODS

Component-level low and high estimates found in STATSGO2 were aggregated to make low and high estimates at the map unit level. Low estimates for an entire map unit were derived by taking the minimum of all the components' low estimates within that map unit. Similarly, high estimates for a map unit were taken to be the maximum of all the components' high estimates within that map unit. The primary interest of this paper is to empirically evaluate the coverage of these PIs using external validation points from across the US.

Soil measurements from an external dataset were used to validate the PI coverage. These validation data were obtained from a 2007 snapshot of the National Soil Characterization Database (National Cooperative Soil Survey 2007) and will be referred to as validation pedons or validation points. The location of all the validation pedons available for the surface horizon (0 to 5 cm) are shown for organic carbon (Fig. 1) and clay content (Fig. 2). The number of validation pedons with attributed measurements varies by soil property. Of the three soil properties considered, organic carbon was generally the least attributed and percent clay was the most attributed.

Because the validation pedon data provide measurements for each horizon at the depths that they actually occupy in the field, these depths had to be converted to the six standard depths used by

GlobalSoilMap. We opted for a depth-weighted average approach (instead of the equal area spline function) in order to keep the validation measurements as representative as possible when they were translated into their corresponding standard depths. After the standard depth property values were calculated for each validation pedon, these values were used to assess the validity of our PI estimates that were obtained using STATSGO2.

One of our primary interests was to empirically evaluate the dependency of PI coverage on other soil characteristics. Specifically, we selected soil order, temperature regime, and moisture regime, which we collectively label as soil classifications throughout this paper. Spatial correlation of the PI coverage was also investigated.

PI coverage was analyzed separately for each combination of depth (six standard depths), soil property (organic carbon, pH, and clay content), and soil classification (soil order, moisture regime, and temperature regime). Pearson's chi-squared tests for independence (Cochran 1952) were used to test if the coverage probability was independent of soil classification.

Moran's I statistic was used to test for spatial patterns of coverage of the validation points. Each point was labeled as falling below, above, or within the associated PI. Moran's I statistic is a global statistical test used to determine if the pattern of coverage exhibits dispersion, complete spatial randomness, or clustering (Bivand et al. 2008).

3 RESULTS

The characteristics of the PIs for organic carbon, pH, and clay are summarized in Table 1. These summary statistics reflect the natural variability of soil properties across the US. The sample medians of the PI widths calculated over the entire US were 1.77% for organic carbon, 2.86 units of pH, and 31.3% for clay content (Table 1).

Across all depth intervals, the proportion of all the validation points that were covered by the corresponding PI was 87.6% for organic carbon, 90.6% for pH, and 86.4% for clay. When summarized by standard depth and by soil classifications, the coverage of the PIs varies from 50% to 100% (Tables 2-10).

Table 1. Five-number summary for PI width.

	Carbon	pH	Clay
Minimum	0.0%	0.36	0.0%
Lower quartile	0.83%	2.21	21.6%
Median	1.77%	2.86	31.3%
Upper quartile	3.20%	3.46	45.3%
Maximum	75.60%	7.16	100.0%

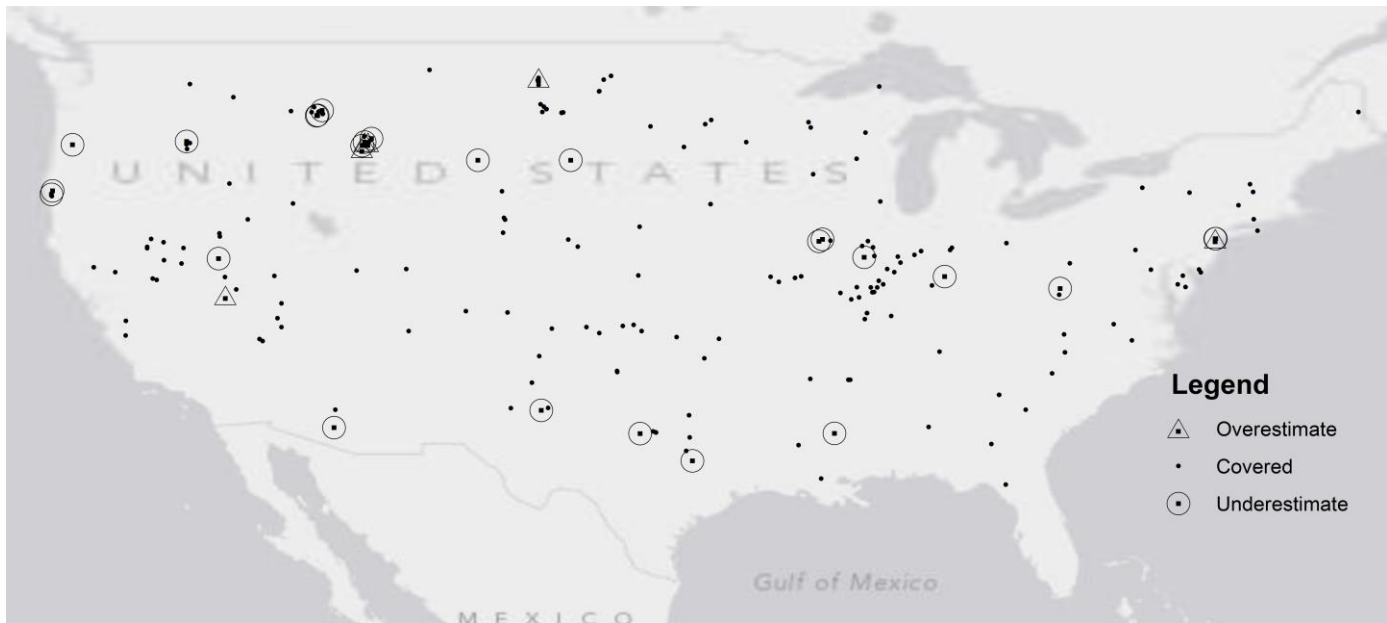


Figure 1: Location of 0 to 5 cm validation points for organic carbon across the conterminous US. Validation pedons with property values that are below the lower prediction limit are highlighted with triangles. Those exceeding the upper prediction limit are highlighted with circles.

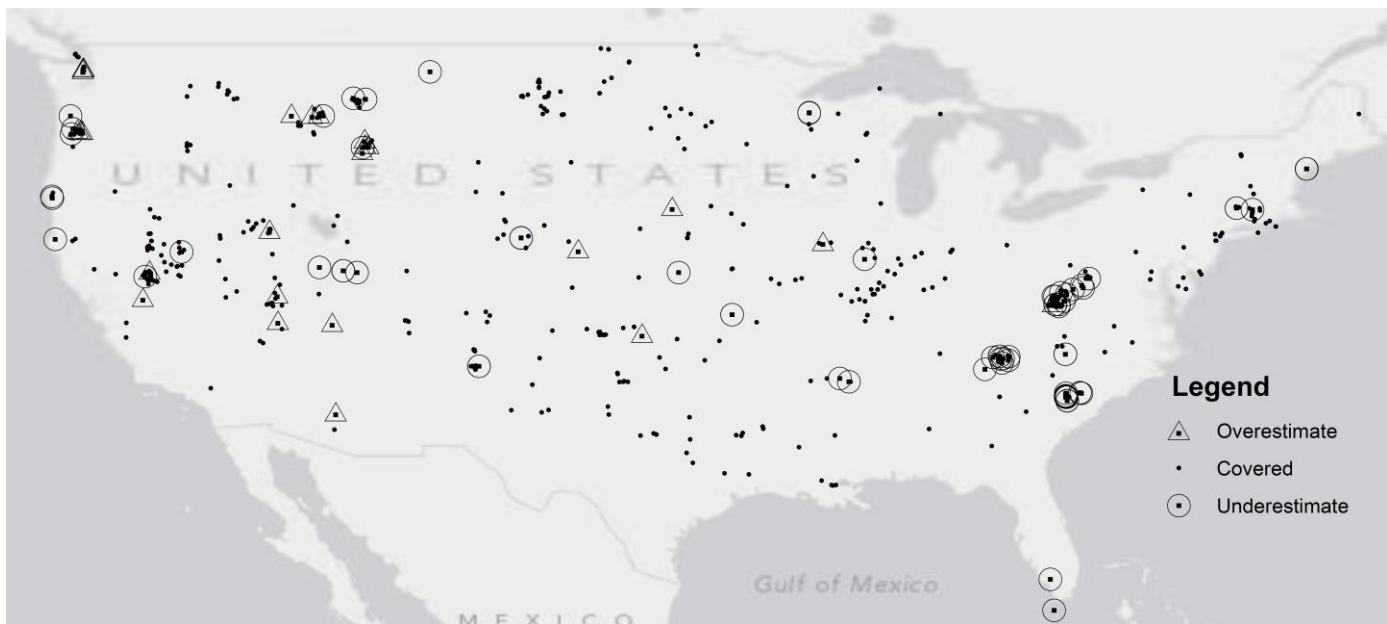


Figure 2: Location of 0 to 5 cm validation points for clay content across the conterminous US. Validation pedons with property values that are below the lower prediction limit are highlighted with triangles. Those exceeding the upper prediction limit are highlighted with circles.

Table 2: Organic carbon coverage chi-squared test for independence: Soil order.

Order*	0-5 cm		5-15 cm		15-30 cm		30-60 cm		60-100 cm		100-200 cm	
	\hat{p}	n	\hat{p}	n	\hat{p}	n	\hat{p}	n	\hat{p}	n	\hat{p}	n
Alfisols	0.925	67	0.938	64	0.930	57	0.912	57	0.887	53	0.907	54
Aridisols	0.870	23	0.950	20	0.955	22	0.850	20	0.923	13	0.889	18
Entisols	0.857	21	0.778	18	0.636	11	0.529	17	0.688	16	0.923	13
Inceptisols	0.571	21	0.550	20	0.500	14	0.750	16	0.909	11	0.778	9
Mollisols	0.906	64	0.910	67	0.933	60	0.932	59	0.893	56	0.863	51
Ultisols	0.963	27	1.000	26	0.956	23	1.000	18	0.895	19	1.000	17
Vertisols	0.833	6	1.000	6	1.000	3	1.000	5	1.000	7	0.857	7
p-value	.0011		.0000		.0000		.0002		.3176		.6591	

* Andisols and Spodosols were removed because of small sample sizes.

Table 3: Organic carbon coverage chi-squared test for independence: Temperature regime.

Temperature regime*	0-5 cm		5-15 cm		15-30 cm		30-60 cm		60-100 cm		100-200 cm	
	\hat{p}	n	\hat{p}	n	\hat{p}	n	\hat{p}	n	\hat{p}	n	\hat{p}	n
Cryic	0.545	11	0.545	11	0.500	8	0.750	8	0.600	5	0.600	5
Frigid	0.956	45	0.936	47	0.977	43	0.955	44	0.974	39	0.838	37
Mesic	0.882	110	0.905	105	0.944	90	0.924	92	0.900	90	0.943	88
p-value	.0007		.0002		.0001		.0709		.0398		.0419	

* Hyperthermic, isomesic, and isothermic were removed because of small sample sizes.

Table 4: Organic carbon coverage chi-squared test for independence: Moisture regime.

Moisture regime	0-5 cm		5-15 cm		15-30 cm		30-60 cm		60-100 cm		100-200 cm	
	\hat{p}	n	\hat{p}	n	\hat{p}	n	\hat{p}	n	\hat{p}	n	\hat{p}	n
Aquic	0.870	23	0.909	22	0.944	18	0.947	19	0.882	17	0.889	18
Aridic	0.880	25	0.957	23	0.958	24	0.833	24	0.889	18	0.952	21
Udic	0.888	98	0.884	95	0.909	77	0.909	77	0.867	75	0.924	66
Ustic	0.889	54	0.963	54	0.940	50	0.922	51	0.959	49	0.846	52
Xeric	1.000	14	1.000	14	1.000	10	1.000	9	1.000	7	0.833	6
<i>p</i> -value	.7571		.3062		.7826		.5492		.4348		.5708	

Table 5: pH coverage chi-squared test for independence: Soil order.

Order*	0-5 cm		5-15 cm		15-30 cm		30-60 cm		60-100 cm		100-200 cm	
	\hat{p}	n	\hat{p}	n	\hat{p}	n	\hat{p}	n	\hat{p}	n	\hat{p}	n
Alfisols	0.880	50	0.889	81	0.957	94	0.967	91	0.904	73	0.857	49
Andisols	0.500	4	0.900	10	1.000	9	1.000	16	1.000	12	1.000	11
Aridisols	1.000	60	0.945	55	0.839	56	0.923	39	0.829	35	0.850	20
Entisols	0.750	44	0.829	47	0.778	36	0.822	45	0.862	29	0.938	16
Inceptisols	0.892	111	0.907	140	0.926	135	0.945	128	0.946	74	0.891	46
Mollisols	0.912	148	0.922	153	0.953	148	0.917	121	0.903	72	0.778	45
Spodosols	1.000	5	1.000	6	1.000	4	1.000	7	1.000	5	1.000	3
Ultisols	0.850	40	0.897	78	0.919	74	0.890	73	0.872	47	0.891	46
Vertisols	1.000	14	0.929	14	1.000	9	1.000	9	1.000	8	1.000	10
<i>p</i> -value	.0006		.6814		.0065		.0793		.4738		.4038	

* Histosols were removed because of small sample size.

Table 6: pH coverage chi-squared test for independence: Temperature regime.

Temperature regime*	0-5 cm		5-15 cm		15-30 cm		30-60 cm		60-100 cm		100-200 cm	
	\hat{p}	n	\hat{p}	n	\hat{p}	n	\hat{p}	n	\hat{p}	n	\hat{p}	n
Cryic	0.793	29	0.818	33	0.818	22	0.941	17	0.933	15	0.875	8
Frigid	0.903	124	0.936	140	0.958	118	0.964	111	0.933	60	0.941	34
Mesic	0.912	216	0.916	297	0.933	315	0.934	303	0.894	207	0.857	147
Thermic	0.951	82	0.956	90	0.899	89	0.875	80	0.919	62	0.854	48
<i>p</i> -value	.0875		.0775		.0782		.1170		.7623		.6073	

* Hyperthermic, isomesic, and isothermic soils were removed because of small sample sizes.

Table 7: pH coverage chi-squared test for independence: Moisture regime.

Moisture regime*	0-5 cm		5-15 cm		15-30 cm		30-60 cm		60-100 cm		100-200 cm	
	\hat{p}	n	\hat{p}	n	\hat{p}	n	\hat{p}	n	\hat{p}	n	\hat{p}	n
Aquic	0.972	36	0.925	40	0.953	43	0.907	43	0.939	33	0.933	15
Aridic	0.959	73	0.913	69	0.838	68	0.920	50	0.872	47	0.846	26
Udic	0.891	138	0.934	226	0.939	247	0.951	245	0.919	149	0.897	107
Ustic	0.901	101	0.940	100	0.929	85	0.917	72	0.898	49	0.795	44
Xeric	0.924	79	0.902	92	0.962	78	0.926	81	0.915	47	0.833	36
<i>p</i> -value	.3124		.8425		.0324		.6849		.8382		.4429	

Table 8: Percent clay coverage chi-squared test for independence: Soil order.

Order*	0-5 cm		5-15 cm		15-30 cm		30-60 cm		60-100 cm		100-200 cm	
	\hat{p}	n	\hat{p}	n	\hat{p}	\hat{p}	n		\hat{p}	n	\hat{p}	n
Alfisols	0.852	122	0.844	122	0.893	112	0.923	117	0.881	109	0.854	103
Andisols	0.800	10	0.882	17	0.867	15	0.812	16	0.857	14	0.917	12
Aridisols	0.915	59	0.889	54	0.915	59	0.912	57	0.857	42	0.932	44
Entisols	0.830	53	0.810	58	0.761	46	0.904	52	0.955	44	0.971	35
Inceptisols	0.712	146	0.767	176	0.834	169	0.902	164	0.942	139	0.936	110
Mollisols	0.961	181	0.941	187	0.947	169	0.912	171	0.878	147	0.877	138
Spodosols	1.000	6	1.000	6	0.800	5	1.000	8	1.000	7	1.000	6
Ultisols	0.785	79	0.773	88	0.835	79	0.904	83	0.872	78	0.847	59
Vertisols	0.786	14	0.714	14	0.778	9	0.818	11	0.939	14	0.786	14
<i>p</i> -value	.0000		.0003		.0131		.8532		.4173		.2126	

Table 9: Percent clay coverage chi-squared test for independence: Temperature regime.

Temperature regime*	0-5 cm		5-15 cm		15-30 cm		30-60 cm		60-100 cm		100-200 cm	
	\hat{p}	n	\hat{p}	n	\hat{p}	n	\hat{p}	n	\hat{p}	n	\hat{p}	n
Cryic	0.789	33	0.824	34	0.857	28	0.900	30	0.864	22	0.941	17
Frigid	0.889	144	0.883	163	0.907	151	0.874	159	0.860	129	0.954	110
Hyperthermic	0.714	7	0.600	5	1.000	5	1.000	6	1.000	5	1.000	5
Mesic	0.876	362	0.858	394	0.880	367	0.914	374	0.890	336	0.850	286
Thermic	0.740	104	0.752	105	0.830	94	0.925	93	0.977	88	0.936	94
<i>p</i> -value	.0038		.0206		.3931		.5206		.0614		.0124	

* Isomesic and isothermic soils were removed because of small sample sizes.

Table 10: Percent clay coverage chi-squared test for independence: Moisture regime.

Moisture regime*	0-5 cm		5-15 cm		15-30 cm		30-60 cm		60-100 cm		100-200 cm	
	\hat{p}	n	\hat{p}	n	\hat{p}	n	\hat{p}	n	\hat{p}	n	\hat{p}	n
Aquic	0.767	60	0.797	59	0.849	53	0.891	55	0.922	51	0.945	55
Aridic	0.875	72	0.853	68	0.859	71	0.884	69	0.852	54	0.926	54
Udic	0.822	276	0.826	322	0.869	297	0.934	302	0.929	269	0.908	217
Ustic	0.930	115	0.915	117	0.935	108	0.883	111	0.906	106	0.922	102
Xeric	0.903	93	0.878	98	0.906	85	0.871	93	0.781	73	0.710	62
<i>p</i> -value	.0090		.1283		.3043		.2554		.0041		.0000	

Results of the chi-squared test for independence can also be found in Tables 2 through 10. Columns represent different standard depths and give both the number of pedons belonging to that specific group (n) and the proportion of these validation points that were covered by the corresponding PI (\hat{p}). At the bottom of each depth column, a *p*-value for the chi-squared test for independence is presented. Because multiple statistical hypothesis tests are being performed, a Bonferroni corrected significance level was used to control the familywise error rate. A family was taken to be the six tests on each of the standard depth intervals that belong the same soil property and soil classification, so a corrected significance level of $\alpha_c = (0.05/6) = 0.008$ was set *a priori*. The individual *p*-values in Tables 2 through 10 should be compared to α_c .

To illustrate how to read the table, the organic carbon 0 to 5 cm standard depth was tested for independence between coverage probability and soil order (Table 2). Seven different soil orders (Alfisols, Aridisols, Entisols, Inceptisols, Mollisols, Ultisols, and Vertisols) were well represented in the validation data giving a total sample size of 229 for this property at this depth. Andisols and Spodosols were removed from this analysis because of low sample sizes ($n < 5$). The chi-squared test has a *p*-value of 0.0011, which indicates that the coverage level depends on soil order. Looking closer, it appears that Inceptisols exhibit lower PI coverage (57.1%) in the validation sample, while Ultisols show a higher coverage (96.3%). Looking at the deeper standard depths, we see this trend continues down through the 30 to 60 cm depth range. However, at the deepest depths, 60 to 100 cm and 100 to 200 cm, this trend dissipates and the corresponding *p*-values (0.3176 and 0.6591, respectively) indicate that the coverage probability does not depend on the soil order at these lower depths. Examining the organic carbon data by temperature regime (Table 3), we see that a dependence exists between temperature regime and coverage probability at the shallowest three depth ranges down to 30 cm; however, at the deeper three depth ranges, this dependence disappears. Similar results for all depths, soil properties, and soil classifications can be found in the remaining tables. Coverage probability is typically between 80% and 95%. In nearly all cases, the coverage probability was independent of the moisture regime, which is important in determining the nationwide consistency of PIs.

Soil pH (Tables 5 through 7) had more validation points than organic carbon. Concerning pH, soil order and coverage probability were dependent at the shallowest depths, but were independent at the deeper depths. For pH, temperature regime and PI coverage were independent at all depths.

For percent clay (Tables 8 through 10), it can be seen that the number of validation points is similar to that of pH. Clay content analysis shows similar results to pH when looking at soil order. The shallower depths, 0 to 5 cm and 5 to 15 cm, show a dependence between soil order and coverage probability, but this dependence disappears as depth increases. A dependence between coverage probability and the five examined temperature regimes was found at the shallowest depth, but not at any of the deeper depths.

Spatial autocorrelation tests using Moran's I statistic reveal the coverage pattern is clustered at nearly all depths for organic carbon, pH, and percent clay (data not shown). The only exceptions, in which no significant spatial autocorrelation was observed, are the organic carbon data for the 0 to 5 cm and 60 to 100 cm depth ranges.

4 DISCUSSION

Developing uncertainty estimates for regional to continental scale maps of soil properties derived from legacy databases can be challenging because of sparse data on soil properties and insufficient information on soil property variability associated with traditional soil surveys (Lilburne et al. 2006). Our approach to developing PIs utilized the estimated data available in the STATSGO2 database to construct map unit level PIs. Given the relatively limited number of sample points with measured soil property data, this approach provides estimates of the lower and upper limits for selected soil properties at any depth and any location with reasonable confidence. While the derived PIs seem to encompass a reasonable proportion of the validation pedon data, the actual widths of these intervals might be larger than is practically useful. Further analysis of the actual PI widths is needed to better assess this.

In general the PIs exhibit more dependence on soil order and temperature regime at shallower depths (0 to 30 cm). This dependence tends to disappear below 30 cm. We posit that this trend may be

the result of the spline algorithm causing systematic shifts of the PIs at shallower depths of soil profiles for certain soil orders and temperature regimes. Other possible explanations could include the greater variability within the surface horizons or an unrepresentative (non-random) selection of validation pedons. All of these possible explanations should be investigated further.

PI coverage almost exclusively appears independent of moisture regime at all depths, which bodes well for stating the consistency of the prediction intervals across the generally gradient of soil moisture regimes in the US. However, the temperature regime and soil order dependencies might suggest latitudinal or elevation difference in data collection methods or overall success of the original mappers in fully characterizing soil property variability.

Spatial autocorrelation tests show clustering at nearly all depth ranges for all soil properties. This is not surprising given the closeness of some of the validation pedons. Intuitively, pedons that are spatially close to one another are likely to share coverage status (i.e. underestimated, covered, or overestimated). The observed spatial dependence between validation points threatens assumptions made by the chi-square tests for independence, but we believe lack of spatial dependence would not change the main findings of this study. To eliminate spatial autocorrelation among our validation points, a more spatially diverse set of sample points selected across the entire United States would likely be beneficial. Although the validation pedons used in this evaluation represent a variety of soil orders, temperature regimes, and moisture regimes from across the conterminous US, they represent points available in a public database and were not the consequence of random selection. However, the aforementioned variety of soil used for validation leads us to believe that the data are adequate to assess the coverage of the PIs and that the results are not heavily influenced by sample selection bias.

The method proposed in this paper for PI creation when legacy data is available has several advantages. First, representative property estimates are not required. Second, the method is simple, which allows it to be easily understood and implemented. Third, for the three soil properties discussed, the observed PI coverage closely approximates the 95% PI coverage targeted by the *GlobalSoilMap* specifications.

As with any method, limitations exist for our proposed method as well. First, PI development requires low and high estimates to be provided at the component level. Second, our method cannot be used to obtain a PI with an adjustable coverage probability (75% for instance).

An important issue that arises is whether the PIs created from this method are too wide for practical use. The reported PI width medians (carbon = 1.77%, pH = 2.86, clay = 31.3%) could be crafted to

encompass a large portion of soils worldwide (e.g. soils with organic carbon between 0.1-1.9, pH 4.0-7.0, and clay 5% to 35%). This highlights the need for more narrow prediction limits. To get narrower PIs, it is likely necessary to begin with a map that does not aggregate up to 21 components per map unit. Applying these same methods to more detailed soil maps, such as the USDA-NRCS Soil Survey Geographic (SSURGO) database, or modifying these methods for use with disaggregated soil survey products (Häring et al. 2012, Nauman & Thompson, accepted) could alleviate this issue.

Even though the collection methods of the legacy data used to derive our PIs is primarily based on expert knowledge rather than objective measurement, we have demonstrated that these data can be used to create PIs that are remarkably consistent across nearly all standard depths, soil orders, moisture regimes, and temperature regimes. In almost all cases, the observed PI coverage is only slightly below the nominal level of 95% and is similar to the empirically derived PI coverage of 91% and 93% reported by Malone et al. (2011). We believe that these PIs derived from the STATSGO2 database represent a reasonable initial product that can serve as a starting point for creating subsequent versions of PIs for *GlobalSoilMap* products for the US.

REFERENCES

- Bivand, R. S., Pebesma, E. J., & Gomez-Rubio, V. 2008. Applied spatial data analysis with R. New York, New York.
- Cochran, W. G. 1952. The χ^2 test of goodness of fit. *Annals of Mathematical Statistics* 25, 315–345.
- GlobalSoilMap Science Committee 2013. *GlobalSoilMap* specifications – Tiered *GlobalSoilMap* products. Release 2.3.
- Häring, T., Dietz, E., Osenstetter, S., Koschitzki, T. & Schröder, B. 2012. Spatial disaggregation of complex soil map units: A decision-tree based approach in Bavarian forest soils. *Geoderma* 185-186, 37–47.
- Lilburne, L., Hewitt, A., & Ferriss, S. 2006. Progress with the design of a soil uncertainty database, and associated tools for simulating spatial realisations of soil properties. In: M. Caetano and M. Painho (Eds) 7th International Symposium on Spatial Accuracy Assessment in Natural Resources and Environmental Sciences, 510–519.
- Malone, B. P., McBratney, A. B. & Minasny, B. 2011. Empirical estimates of uncertainty for mapping continuous depth functions of soil attributes. *Geoderma* 160(3), 614–626.
- National Cooperative Soil Survey. 2007. National cooperative soil characterization database. <http://ncsslabdatamart.sc.egov.usda.gov>.
- Nauman, T. W. & Thompson, J. A. Semi-automated disaggregation of conventional soil maps using knowledge driven data mining and classification trees. *Geoderma*, Accepted.
- Ogders, N. P., Libohova, Z. L., & Thompson, J. A. 2012. Equal-area spline functions applied to a legacy soil database to create weighted-means maps of soil organic carbon at a continental scale. *Geoderma* 189-190, 153–163.
- USDA-NRCS 2011. U.S. general soil map (STATSGO2) description. http://soils.usda.gov/survey/geography/ssurgo/description_statsgo2.html.